

ORIGINAL RESEARCH PAPER

Domestic smart metering infrastructure and a method for home appliances identification using low-rate power consumption data

Ioannis Paraskevas¹  | Maria Barbarosou² | Richard Fitton¹ | William Swan¹

¹Smart Meters > Smart Homes Lab, Energy House Laboratories, University of Salford, Manchester, UK

²Department of Aeronautical Science, Division of Electronics, Electric Power and Telecommunications, Hellenic Air Force Academy, Athens, Greece

Correspondence

Ioannis Paraskevas, Smart Meters > Smart Homes (SMSh) Lab, Joule House, University of Salford, Manchester M5 4WT, UK.
Email: I.Paraskevas@salford.ac.uk

Abstract

The deployment of domestic smart metering infrastructure in Great Britain provides the opportunity for identification of home appliances utilising non-intrusive load monitoring methods. Identifying the energy consumption of certain home appliances generates useful insights for the energy suppliers and for other bodies with a vested interest in energy consumption. Consequently, the domestic smart metering system, which is an integral part of the smart cities' infrastructure, can also be used for home appliance identification purposes taking into account the limitations of the system. In this article, a step-by-step description on accessing data directly from the domestic Smart Meter via an external Consumer Access Device is described, as well as an easy-to-implement method for identifying commonly used home appliances through their power consumption signals sampled at a rate similar to the rate available by the domestic smart metering system. The experimental results indicate that the combination of time domain with frequency domain features extracted either from the 1D/2D Discrete Fourier Transform or the Discrete Cosine Transform provides improved recognition performance compared to the case where the time domain or the frequency domain features are used separately.

1 | INTRODUCTION

The aims of the domestic smart metering rollout in Great Britain were to enable consumers to monitor and control their energy usage and to assist in the transition of the country towards a low-carbon economy [1]. The initial target of the UK government was to install smart metering systems in all houses and small businesses in Great Britain by the end of 2020 [2]. However, this deadline has now been extended to 2024 [3, 4]. Data from approximately 21 million smart metering systems, installed in Great Britain [3], are gathered by the energy supply companies for both SMETS1 and SMETS2. SMETS stands for Smart Metering Equipment Technical Specifications; SMETS2 systems have advanced features compared to SMETS1 [5, 6].

The deployment of the domestic smart metering infrastructure has already contributed towards the introduction of new products and technologies in the energy market such as the dynamic Time of Use (ToU) tariffs which aim towards shifting the consumers' higher energy demands to off-peak periods, thus, contributing to the stability of the grid [7]. Grid stability will become even more imperative with the

gradual increase in the number of electric vehicles in the market, thus, monitoring the consumers' energy consumption is crucial. The motivation for this work is the utilisation of the domestic smart metering infrastructure to develop a useful application using power consumption analytics [7]. Specifically, to develop an end-to-end solution for home appliance identification which would be beneficial for both the consumer and the energy supply company thus, harnessing the benefits of the data generated from the domestic smart metering system.

The standard domestic smart metering system consists of three devices: the electricity meter, the gas meter and the In-Home Display (IHD). The IHD provides the consumer with near real-time information about their electricity and gas energy consumption, the corresponding cost, and the amount of CO₂ produced due to the energy usage alongside additional energy-related information.

The data generated by each domestic smart metering system in Great Britain is gathered by the Data Communications Company (DCC) [8] and is shared with the energy supply company (Figure 1, taken from [9]) primarily for billing purposes. However, consumption data, can also be used for

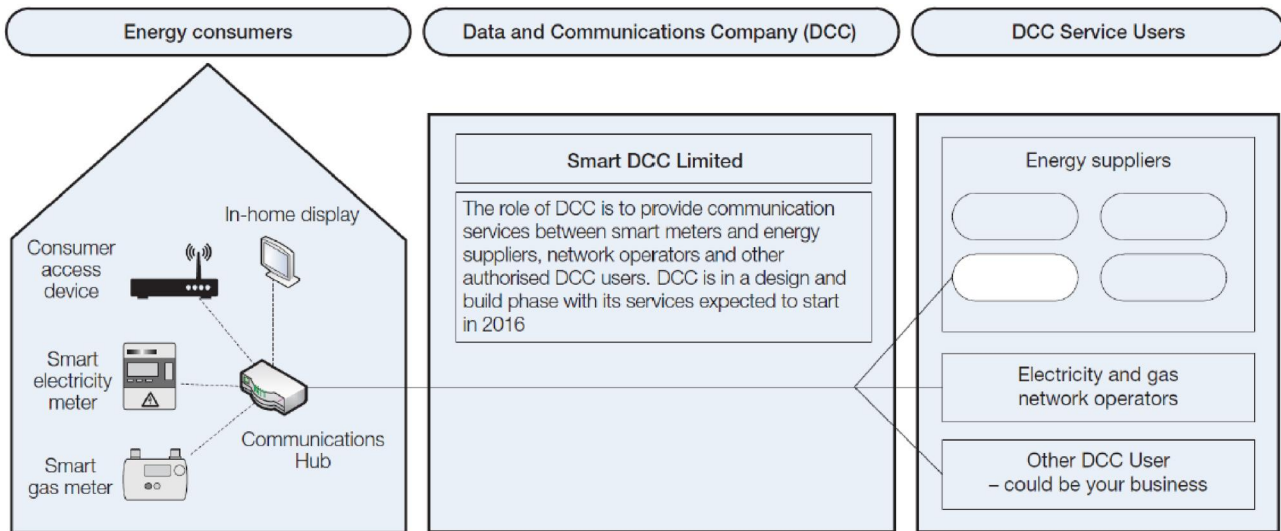


FIGURE 1 Overview of the end-to-end system

monitoring purposes, providing insights to customers' energy consumption patterns as well as for identifying the use of certain domestic appliances to provide consumers with energy and cost-saving advice.

It should be noted that the current domestic smart metering infrastructure provides the DCC and the energy supplier with direct access to the consumer's smart metering data (Figure 1); the consumers can access their data indirectly through the energy supplier with a 30-min granularity though. In order for the consumer to access their data directly and with granularity higher than 30 min, additional hardware is required. Thus, recently, 'external' Consumer Access Devices (CADs) have been introduced to the market [9, 10]. External CADs are paired with the domestic smart metering system in order to provide the customer with access to their energy-related data via the cloud [10]. To avoid confusion, the domestic Electricity and Gas Smart Meters also incorporate a CAD ('internal' CAD) in order to transmit the measured quantities to the DCC/energy supplier however, the 'external' CAD is an additional hardware device which is paired with the domestic smart metering infrastructure so that the consumer can access their energy-related data directly. External CADs usually use the Representational State Transfer (REST) Application Programming Interface (API) and/or the MQ Telemetry Transport (MQTT) API [11]. Consequently, the consumers, similar to the energy suppliers, can access and download their energy and power consumption data for near-real-time monitoring as well as for a variety of applications. The highest data granularity that the domestic smart metering infrastructure can provide through either the internal or external CAD is 10 s which is an important parameter when designing the related algorithms [5].

In brief, taking into consideration on one hand the limitations of the domestic smart metering data in terms of granularity and on the other hand the opportunity for the consumer to engage with their data, this work aims to: (i) provide a description on how the consumers can access their data from the

domestic smart metering system via the MQTT API through the external CAD and (ii) present and test an easy-to-implement method for home appliance identification by utilising power consumption data with granularity similar to the granularity available by the domestic smart metering system which is 10 s. The experimental results of the proposed method demonstrate the potential of utilising the data acquired from the smart meters not only for applications related to billing, grid stability etc., but also for classification-related applications.

Thus, the main contribution of this work is the development of an end-to-end comprehensive solution for home appliance identification at no or very low cost utilising the domestic smart metering infrastructure. Specifically, the electricity smart meter is paired with an external CAD (provided freely by certain Energy Suppliers or purchased at a low cost) enabling the consumers to access their power consumption data from the CAD's cloud through a light network protocol such as the MQTT. A novel algorithm is implemented and tested for the home appliance identification task; similar algorithms have also been used in the field of audio/speech pattern recognition which, like the power consumption signals, are one-dimensional [12]. For the proposed method, features from the time domain are combined with features from the frequency domain in order to identify a number of home appliances which are common in almost every household in the UK. The largest (peak) values are selected from the time domain of the power consumption signal and are combined with coefficients from the frequency domain. The time domain features (power values) and the frequency domain features are then introduced to an easy-to-implement single hidden layer Feedforward Neural Network (FFNN). In brief, the experimental results show that the use of the time domain features provides higher identification rates compared to the rates reached using the frequency domain features. However, the highest identification score is achieved when features from both the time and the frequency domains are utilised.

Specifically, for certain home appliances (classes) where the time domain features underperform, the frequency domain coefficients compensate for this low performance, indicating that the two feature sets complement each other.

In order to test the proposed method, 6-s and 12-s power consumption data from the UK Domestic Appliance-Level Electricity (UK-DALE) dataset, which is an annotated publicly available dataset [13, 14], is used to identify eight commonly used home appliances utilising 24-h signal signatures for the identification of each device. UK-DALE provides 6-s data so, the 12-s data was derived from the 6-s dataset by selecting every other sample. The experimental results acquired through the 12-s signals were particularly important because this sampling period is very close to the 10-s sampling period available by the UK domestic smart metering infrastructure.

The rest of the paper is organised as follows: In Section 2, the literature review on the identification of home appliances is provided and in Section 3, a description on how the consumers can access their data through an external CAD is given. Specifically, a brief overview of the API used is provided alongside the steps which need to be followed by the consumers to access their energy/power consumption data through the external CAD's cloud service. In Sections 4 and 5, the proposed method for identifying the home appliances alongside the experimental results is presented and discussed. In Section 6, the conclusions and summary of this work are given.

2 | HOME APPLIANCE IDENTIFICATION USING NON-INTRUSIVE LOAD MONITORING

The disaggregation of the domestic smart metering data, also called non-intrusive load monitoring (NILM), introduced by Hart in [15], aims to identify which appliances are being used by processing the household's aggregated power consumption signal. In general, home appliance disaggregation is achieved following the standard pattern recognition procedure utilising a set of features for identifying each appliance. The selection of the appropriate features as well as the classification algorithms depend, amongst other parameters, on the sampling rate of the aggregated power signal. Generally, features can be extracted during the transient or steady-state operation of an appliance [16, 17]. The features which correspond to the devices' transient state operation require higher sampling frequencies compared to the features extracted during the appliance's steady-state operation. Steady-state features are the root mean square, peak values of the electrical signals, the power factor, power changes, V-I characteristic shape features, steady state signal harmonics, electromagnetic interference signatures etc. Steady-state features can also be extracted from the real and the reactive power. The transient parameters are related to the duration, shape and size of the current's transient waveform, the current signal spikes, the response time etc. More generally, the features used for NILM could be categorized as microscopic or 'high frequency' and macroscopic or 'low frequency' [17, 18]. Common macroscopic features are the variations of

the real and the reactive power signals, power quality indicators, and the temporal discrete power pulses. The microscopic features are extracted from the frequency domain of the signal and could be the harmonics of the electrical signals, from the noise spectrum of the voltage signal, from the signal's wavelet transform etc.

It is important to mention that the large-scale deployment of the domestic smart metering systems over the last decade has shifted the focus of research from the analysis of higher sampling rate meter readings (kHz to MHz range) to lower rate meter readings (sampled at 1–60 s) as well as to very low-rate meter readings in the 15–60 min range [19].

It is important to underline that the 10-s power consumption signal of the domestic smart metering system (SMETS1/SMETS2) is the real power and thus, features from the reactive and apparent power, the power factor etc., are not available [5]. Consequently, as mentioned earlier, for the proposed method, time domain features as well as coefficients from the frequency domain of the real power consumption signal are utilised. Then, an easy-to-implement Neural Network is used for home appliance identification. The easy-to-implement FFNN is one of the attributes which differentiates this work compared to the majority of the recently published research in this field where deep learning (DL) models, namely deep neural networks (DNNs) and convolutional neural networks (CNNs), are utilised resulting in an increased computational cost.

In the remaining part of this section, a review of published work in the field of NILM is provided starting from two research works utilising signals sampled at a high rate, then continuing with publications using signals sampled at a low rate, which is the primary focus of this work, and closing with published work using signals of very low rate.

In both [20, 21], two event-based NILM classification algorithms using high sampled current data are presented. In both works, image-like representations of the signals are developed and then introduced to a CNN for the classification task. Specifically, from the one-cycle activation current of each appliance a weighted recurrence graph is developed in [20] and the Fryze power theory is used in order to decompose it into its active and non-active components and, subsequently, the 2D Euclidean-distance similarity matrix is used to represent the decomposed current signal into an image [21]. The methods presented in [20, 21] are evaluated using the PLAID dataset, which contains measurements sampled at 30 kHz and the method in [20] is also tested using LILACD which is an industrial dataset with three phase data sampled at 50 kHz. In both works, the recognition rate reached was very high.

A real-time NILM algorithm was introduced in [22]. In this work, a super-state hidden Markov model and a sparse Viterbi algorithm were developed for disaggregating low and very low frequency data. The proposed algorithm is appropriate for disaggregating appliances with complex multi-state power signatures. The method was tested on 18 loads from the REDD dataset (1/3 Hz sampling frequency) using the apparent power signal as the feature and on the AMPds dataset (1/60 Hz sampling frequency) using the current signal as the feature with impressive classification accuracy results. Moreover, it is

important to underline the efficiency of the method as it can run on an embedded processor. In [23], a feature set consisting of Mel-Frequency Cepstral Coefficients (MFCCs), Spectrogram and Mel-Spectrogram time-frequency distribution features is introduced to a multi-layer Long Short-Term Memory recurrent neural network (RNN) and in [24], the spike histogram—which is a time-power distribution evaluated by taking the differences between consecutive power data points—of the power consumption signal is introduced to different DL architectures which are compared in terms of their recognition performance. This last work initiated a research collaboration which resulted in building the NILM Toolkit which is an open-source software incorporating free datasets and metrics in order to assist researchers to develop and validate their data disaggregation algorithms [25]. In [26], a Very Deep One-Dimensional (1D) CNN with 13 1D convolutional operations grouped into five classical convolution layers and three fully connected layers was implemented for the application of home appliances' power signature classification. In [27], a three-layer CNN was developed for the same application and seven power signals (classes), sampled at 1 Hz, were introduced to the classifier. In [28], the WaveNet, which is a DL-based architecture, is shown to be better, in terms of handling a long time series, compared to the CNNs and the RNNs for the task of home appliances power disaggregation. Another approach was presented in [29], where a Convolutional Variational Autoencoder, which is a combination of a Variational Autoencoder and a CNN, was used for energy disaggregation and also in [30], a 1D CNN RNN was used for the same task. In [31], the active power and its step change as well as the reactive power were used as features for the home appliance classification task. Four commonly used classifiers namely, Decision Tree, Nearest Neighbor, Discriminant Analysis and a FFNN were utilised. Moreover in [32], 19 features were extracted from the power consumption signals, from the appliances' time usage and their location in the household. A random forest classifier provided the highest classification accuracy utilising the extracted features. In [33],

features from the home appliances' curves were combined with occupant-related behavioural features and the appliances' power range. A Bayes model was then used for the classification of seven appliances.

In [34], 15-min data was acquired from the smart metering system in order to identify the electric resistance water heater load. In the same work, it was highlighted that the 15-min data available was not adequate for identifying the load and 1-min data had to be used instead. Moreover, in [35], 15-min smart metering data alongside weather data were introduced to a random forest classifier for the application of household classification.

3 | ACCESSING SMART METERING DATA

This section focuses on the structure of the smart metering cluster and on a popular API protocol which can be used in order to access smart metering data using an external CAD. Specifically, an overview of the smart metering cluster 0x0702 is provided as well as a step-by-step description on accessing the smart metering data using the MQTT API [36–38].

3.1 | The smart metering cluster

The data acquired from any domestic smart meter is organised into a standard interface in order to ensure interoperability among different devices [37, 39]. This interface is common to the Electricity Smart Metering Equipment (ESME) and the Gas Smart Metering Equipment (GSME), and is based on the standardised 0x0702 smart metering cluster which is organised in four main sections: (i) formatting, (ii) reading information set, (iii) historical consumption and (iv) meter status [40]; (see Table 1).

TABLE 1 Structure of the smart metering cluster 0x0702; ZigBee Alliance standard, smart energy profile

Attribute set identifier	Attributes	Description
Formatting 0x03	Energy consumption and demand format	Deciphers the energy and demand-related attributes
	Multiplier/divisor	Multiplies/divides the received values in order to express them in the selected unit; kWh and kW for the smart metering case
	MPAN	Meter Point Administration Number
	Smart meter serial number	
	Energy measurement unit	For electricity and gas in kWh
	Commodity	Electricity or gas
Reading information set 0x00	Current summation delivered	Current value of energy delivered to the household
	Current maximum demand delivered	
ESP historical consumption 0x04	Current day consumption delivered	
	Instantaneous demand	
Meter status 0x02	Status	

Additional information is also included in the cluster such as the relevant time stamps in Unix and Portable Operating System Interface (POSIX) formats, the Link Quality Indicator (LQI) of the Personal Area Network (PAN) etc.

In the smart metering cluster presented in Table 1, one can observe that the attribute sets which contain dynamic data are the Reading Information 0x00 set and the Energy Service Portal (ESP) Historical Consumption 0x04 set. These two sets contain energy and power-related information such as the total energy received by the household, the energy consumption of the household during the current day and the instantaneous power demand of the household. The 0x03 and 0x02 attribute sets contain static data including data formatting, the serial number of the smart electricity or gas meter, the Meter Point Administration Number (MPAN) etc.

3.2 | Interfaces for accessing the smart metering data

The domestic smart metering data can be accessed by the consumer from the cloud via an external CAD in near real-time. Commonly used interfaces for accessing data from the CAD's Cloud are the REST and the MQTT API [11]. Certain energy supply companies may also provide direct access to the consumer's energy data through their portal although in this case, the data granularity is much lower (30-min) compared with the 10-s data granularity available through the external CAD [41, 42].

The REST API uses the well-known HTTP protocol and a Representation State Transfer software architectural style whereas, the MQTT API uses the TCP/IP protocol and utilises a publish/subscribe architectural style. Moreover, the MQTT API does not require polling the server for receiving the generated data [11]. Thus, by using the MQTT API the consumer can receive their data without the extra step of polling the CAD's cloud server which makes the MQTT more appropriate for the application presented here. In order to process and post-process the power consumption/energy data received from the CAD's cloud server, certain steps need to be taken because the raw data is in hexadecimal notation, and incorporates UNIX timestamps etc. (Figure 2). In brief, the smart metering data can be accessed utilising the MQTT API through a Linux-based Operating System using a version of the following code:

```
mosquitto_sub -h XXX.XX.XXX.XX -u Username -P
Password -t 'XXX/XXX/MAC address'
```

where,

mosquitto_sub, corresponds to the MQTT subscribe process,
h, is the hostname, that is, IP of the server where the data will be accessed from,
u, is the username,
P, is the password,
t, is the message topic which should include the MAC address of the CAD device.

Thus, every 10 s the user will receive the updated measured quantities of the cluster presented in Figure 2; an alternative presentation (organised form) of Figure 2 is presented in Figure 1a (Appendix). The acquired data can then be uploaded to a software package for near real-time processing and post-processing purposes. In both Figures 2 and 1a, the MPAN and the electricity smart meter serial number have been omitted on purpose.

The most important quantity for the proposed home appliance identification method is the Instantaneous (power) demand identified in Table 1 (alongside Figures 2 and 1a); Attribute Set Identifier 0x04. The instantaneous demand quantity needs to be converted from hexadecimal to decimal and then divided by 1000 (3E8 in hexadecimal) as pointed out in the Attribute Set Identifier 0x03 of Table 1 as well as in Figures 2 and 1a [40]. After this process has been completed, the Instantaneous (power) demand will be in kW units.

4 | PROPOSED METHOD FOR HOME APPLIANCE IDENTIFICATION USING THE DCT AND THE DFT

In this work, some of the most commonly used home appliances are identified through their 24-h real power signals, sampled every 6 and 12 s, utilizing features from both the time and frequency domains. The home appliance power signals used for testing the proposed method were taken from the UK-DALE dataset.

For the experiments, the largest power values of the signals were selected from the time domain alongside the largest magnitude and amplitude coefficients from the 1D and the 2D-Discrete Fourier Transform (DFT) and the 1D and the 2D-Discrete Cosine Transform (DCT), respectively, of the power consumption signals and then introduced to a FFNN. The reasons behind using the DCT-based features are as

Message topic/CAD's MAC address

```
{"elecMtr":{"0702":{"03":{"01":"000001","05":"43","04":"9B","02":"0003E8","07":"","08":"","03":"43","00":"00","06":"00"},"07":"233FA48A","01":"000000000000","00":"0000002AF18E","14":"02","02":"000000001D8E"},"04":{"01":"0070E4","00":"000E43","02":{"00":"00"}}},"ts":"2018-09-27 14:24:14","time":"5BACE80E","gmtime":1538058254,"pan":{"status":"joined","nPAN":"00","join":"0","lqi":"FF"},"gts":"2018-09-27 02:00:02","ets":"2018-09-27 14:24:10","error":{"last Command":"00","errorCode":"00"},"gid":""}}
```

FIGURE 2 Smart metering cluster/MQTT API (raw form)

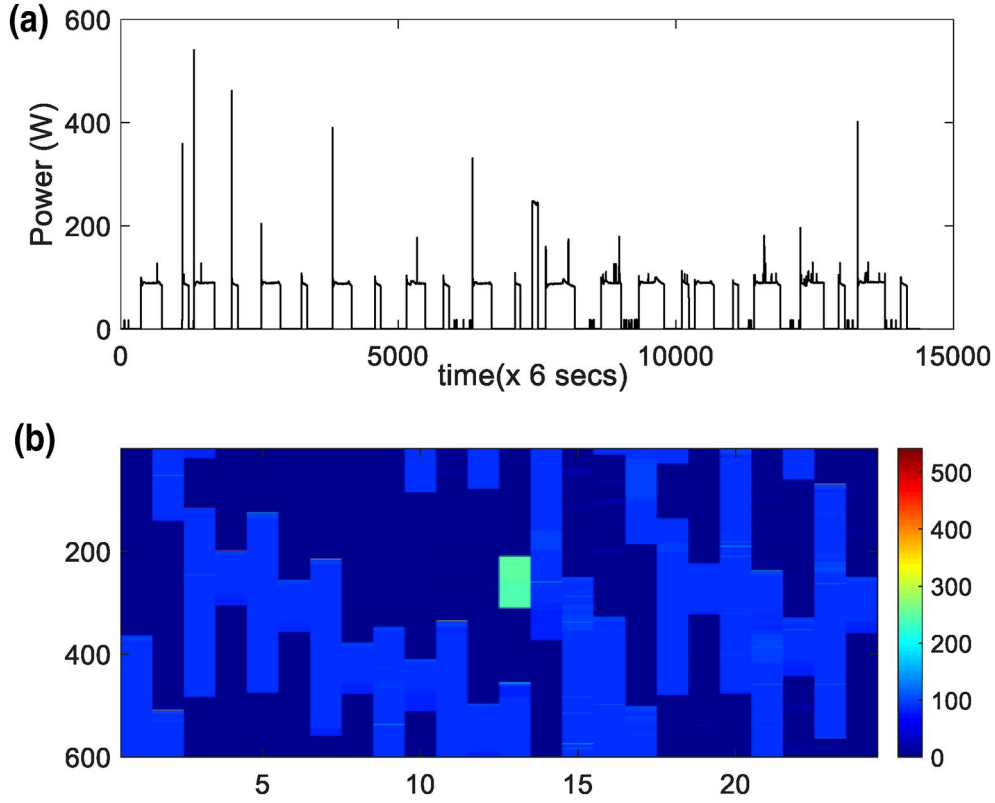


FIGURE 3 Twenty-four hour power signal of fridge: (a) in 1D and (b) reshaped in 2D

follows: (i) DCT is a real transform and consequently only calculations of real numbers are required thus, reducing the computational effort and memory requirement and (ii) DCT is well-known for its energy compaction capability meaning that it can encapsulate most of the signal's energy content in a few coefficients [43]. For this reason, both the commonly used DFT as well as the DCT were tested in terms of their feature extraction capabilities. Furthermore, the 2D versions of the transforms were also tested to investigate whether they are more robust, in terms of their recognition performance, compared to the 1D versions of the same transforms. To apply the 2D transform on the 1D power signal, the original time series needs to be reshaped from 1D to 2D. As an example, a 24 h power signal corresponding to the fridge device in its 1D and 2D representations is depicted in Figure 3a and b, respectively.

Figure 4 shows the resulting distribution (heatmap) and histogram graphs of the 2D-DCT (Figure 4a and b) and the 2D-DFT (Figure 4c and d) of the 2D signal illustrated in Figure 3b. It is apparent from the figures that the high intensity coefficients of the 2D-DCT and 2D-DFT transforms are concentrated in a small area in the upper (Figure 4a) and middle rows (Figure 4c), respectively. This result was expected, since the original signal has dominant vertical stripes, as can be seen in Figure 3b. Note that the area where the largest coefficients are concentrated for the 2D-DCT case (Figure 4a) is smaller compared to the 2D-DFT (Figure 4c), which indicates the DCT's superior compaction capability.

4.1 | DCT and DFT definitions

In this section, the definitions of the 1D and 2D-DCT and DFT are summarised. Given an 1D discrete-time signal $x(n)$ where n corresponds to the time samples with $n = 0, 1, \dots, N-1$ and a 2D discrete-time signal $X(k, l)$ with $k = 0, 1, \dots, M-1$ and $l = 0, 1, \dots, N-1$, the mathematical formulas for the 1D-DCT and DFT and the 2D-DCT and DFT, are as follows [43, 44].

1D-DCT:

$$y(k) = a(k) \sum_{n=0}^{N-1} x(n) \cos \frac{\pi(2n+1)k}{2N}, \quad k = 0, 1, \dots, N-1$$

where

$$a(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k \neq 0 \end{cases}$$

2D-DCT:

$$Y(k, l) = a(k)a(l) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(X(m, n) \cos \frac{\pi(2m+1)k}{2M} \cdot \cos \frac{\pi(2n+1)l}{2N} \right),$$

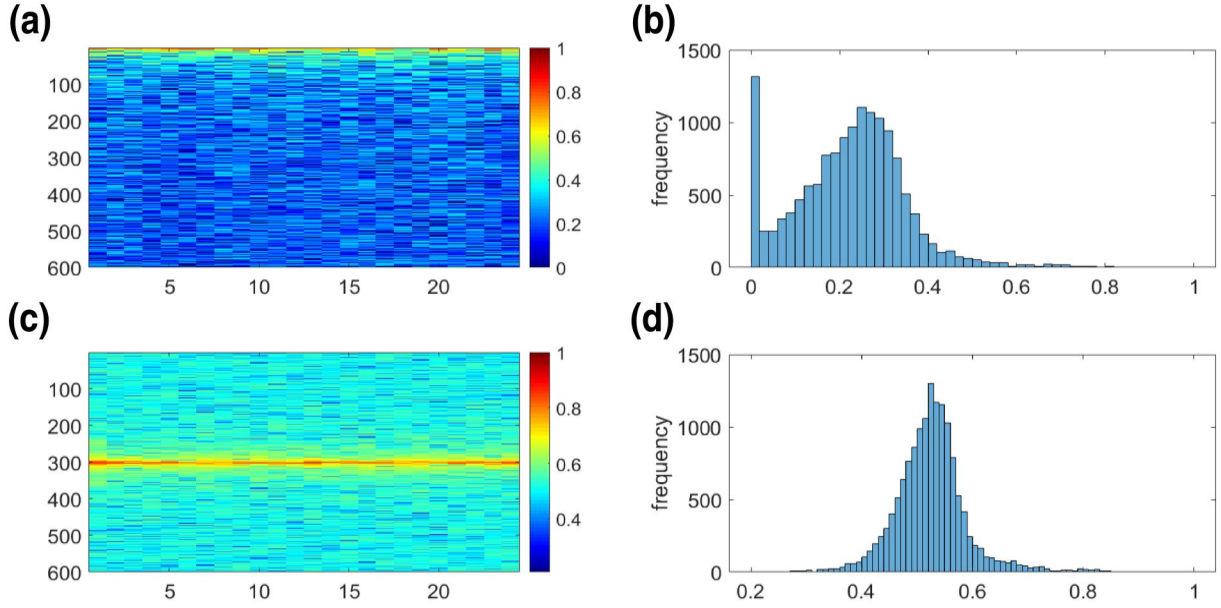


FIGURE 4 Heatmaps and histograms rescaled to $[0,1]$ of the (a and b) 2D-DCT and (c and d) 2D-DFT applied on the 2D 24 h signal illustrated in Figure 3b

$$k = 0, 1, \dots, M-1, l = 0, 1, \dots, N-1$$

where

$$a(k) = \begin{cases} \frac{1}{\sqrt{M}}, & k = 0 \\ \sqrt{\frac{2}{M}}, & k \neq 0 \end{cases} \quad \text{and} \quad a(l) = \begin{cases} \frac{1}{\sqrt{N}}, & l = 0 \\ \sqrt{\frac{2}{N}}, & l \neq 0 \end{cases}$$

1D-DFT:

$$y(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) e^{-j2\pi \frac{nk}{N}} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \cdot \left[\cos\left(2\pi \frac{nk}{N}\right) - j \sin\left(2\pi \frac{nk}{N}\right) \right], \quad k = 0, 1, \dots, N-1$$

and

2D-DFT:

$$Y(k, l) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(m, n) e^{-j2\pi \left(\frac{km}{M} + \frac{ln}{N}\right)},$$

$$k = 0, 1, \dots, M-1, l = 0, 1, \dots, N-1$$

4.2 | Dataset

The proposed method for home appliance identification, which is described in Section 4.3, is tested using the UK-DALE dataset [13, 14]. The UK-DALE dataset was selected,

from among other popular datasets, because it is open, incorporates home appliances which are typical in a wide range of UK households and also, the appliances' power consumption signals have long duration which is necessary for training the FFNN [45]. Specifically, the UK-DALE-2017 version includes power consumption signals of home appliances from 5 UK homes with a 6-s sampling period.

In this study, the data of house 1 is used, which covers a period of 4.3 years starting from 09/11/2012 until 26/04/2017. To test the proposed method, two sets of experiments were conducted. One with signals having 6-s granularity and the other set of experiments using the same signals however, the granularity was decreased to 12-s. The 12-s granularity was obtained by downsampling the 6-s signals by selecting every other sample thus, reducing the sampling frequency to 1/12 Hz. The reason behind conducting experiments with the 12-s sampling period is to simulate a near real-case domestic smart metering scenario because, as mentioned earlier, the highest data granularity acquired from the domestic smart meters is 10 s.

The classes of appliances which were selected in order to test the proposed method are presented in Table 2. These appliances were chosen because they form a group of appliances commonly used in the majority of the UK households.

On purpose, the boiler and the gas oven, which are primarily gas consuming devices, have been incorporated in the group in order to show the potential of the proposed method to identify appliances which are not primarily electrical. For the identification of these gas consuming devices, the power consumption patterns of their electrical components such as ignition, fan etc., have been utilised. The meaning of the parentheses used in the 'No. of 24 h power signals' of each class presented in Table 2 is explained in the following section.

TABLE 2 Home appliances dataset

Class/appliance	No. of 24 h power signals (with 12 h overlapping)	
	Sampling period	
	6 s	12 s
1/Boiler	2877 (2954)	2872 (2954)
2/Washing machine	1942 (2715)	1894 (2715)
3/Dishwasher	1423 (2752)	1413 (2752)
4/TV	2419 (2743)	2415 (2743)
5/Kettle	2386 (2621)	2377 (2621)
6/Fridge	2690 (2690)	2690 (2690)
7/Lighting circuit	2695 (2885)	2695 (2885)
8/Gas oven	2524 (2524)	2510 (2524)

4.3 | Feature sets

The steps followed in order to develop the feature sets which are introduced to the FFNN are summarised as follows.

4.3.1 | Usage of the 1D-DCT and the 1D-DFT

- (i) *Development of the 24 h power consumption series dataset:* Each one of the power consumption time series of the selected appliances (Table 2) is segmented into frames with 24 h duration and 12 h overlapping; each frame consists of 14,400 or 7200 values depending on the sampling rate (1/6 Hz or 1/12 Hz).
- (ii) *Removal of the 24 h signals that correspond to the appliances' off or standby mode:* Following (i), the maximum power value of each 24 h power consumption signal is detected and if its power is less than 3W—which usually corresponds to the consumption of the appliance's remote signal sensors and their circuits, to miscellaneous LED status lights, digital clocks, touch-on button etc.—then, the signal is deleted. In Table 2, the numbers outside the parentheses represent the total number of the 24 h power consumption signals for the 6-s and the 12-s sampling periods for each appliance after steps (i) and (ii) have been completed whereas, the numbers inside the parentheses correspond to the total number of the 24 h power consumption signals before the 3W threshold was applied.
- (iii) *Application of the 1D-DCT and the 1D-DFT on the 24 h power consumption signals:* The 1D-DCT and the 1D-DFT are applied on each one of the 24 h frames.
- (iv) *Feature set:* The feature vector includes: (a) the time domain features which are the T largest power values and (b) the frequency domain features which are either: the F largest absolute value coefficients (real numbers) of the 1D-DCT, or the F largest magnitude value coefficients (complex numbers) of the 1D-DFT from each 24 h signal. In Section 5, the classification scores are presented for the cases where feature vectors of maximum length 5

($T + F = 5$) and 10 ($T + F = 10$) are introduced to the classifier.

4.3.2 | Usage of the 2D-DCT and the 2D-DFT

For the 2D-DCT and the 2D-DFT cases, the 24 h power signals are reshaped from 1D to 2D. In this case, the steps for developing the feature set are organised as follows:

- (i)–(ii) Same as for the 1D case.
- (iii) *Reshape of the power consumption signals from 1D to 2D:* The 24 h power time series are reshaped from vectors to matrices. Specifically, each time series consisting of 14,400 samples (6 s sampling period) is reshaped into a 600×24 matrix and each 7200 samples time series (12-s sampling period) is reshaped into a 600×12 matrix.
- (iv) *Application of the 2D-DCT and the 2D-DFT on the power consumption matrices:* Apply the 2D-DCT and the 2D-DFT on the matrices formed in step (iii).
- (v) *Features set:* Similar to the 1D case, the feature vector includes: (a) the T largest power values (same time domain features as in 4.3.1) and (b) either: the F largest absolute value coefficients of the 2D-DCT or the F largest magnitude value coefficients of the 2D DFT. As for the 1D case, feature vectors of size $T + F = 5$ and $T + F = 10$ are formed and then introduced to the classifier in order to evaluate their recognition performance.

4.4 | Evaluation of the proposed method using the confusion matrix

In this section, a brief description of the confusion matrix concept is provided, because in Section 5 confusion matrices will be utilised to demonstrate the accuracy of the proposed method. The rows of a confusion matrix, for example, in Figure 5 (usually labelled as the Output Class), correspond to the predicted class (output) of the classifier whereas, the columns of the confusion matrix (usually labelled as the Target Class) correspond to the correct (true) class. Any element C_{ij} —where i and j are the row and column of the confusion matrix, respectively—corresponds to the number of power signals which were classified in class i (Output Class) although they belonged to class j (Target Class). The cells on the diagonal of the confusion matrix correspond to the number of power signals which have been correctly classified.

The classification performance metrics Recall, Precision and Overall Accuracy can be directly derived from the confusion matrix. Specifically, Precision, P_k , is the percentage of the correctly classified power signals in class k divided by the total number of power signals classified in this class. The percentages in the cells of the right-most column of the confusion matrix correspond to the Precision values of each class. Recall, R_k , is the percentage of the correctly classified power signals in class k divided by the total number of power signals which belonged to class k . The percentages in the cells

FIGURE 5 Confusion matrix for $T = 10$ and $F = 0$; 12-s granularity (Table 6, first row)

Confusion Matrix									
Output Class	1	2	3	4	5	6	7	8	
	2542 13.5%	3 0.0%	13 0.1%	64 0.3%	1 0.0%	0 0.0%	6 0.0%	4 0.0%	96.5% 3.5%
	0 0.0%	1680 8.9%	15 0.1%	0 0.0%	18 0.1%	9 0.0%	21 0.1%	0 0.0%	96.4% 3.6%
	0 0.0%	5 0.0%	916 4.9%	0 0.0%	140 0.7%	0 0.0%	0 0.0%	0 0.0%	86.3% 13.7%
	17 0.1%	56 0.3%	28 0.1%	2331 12.4%	2 0.0%	0 0.0%	120 0.6%	6 0.0%	91.1% 8.9%
	0 0.0%	28 0.1%	406 2.2%	0 0.0%	2210 11.7%	0 0.0%	0 0.0%	0 0.0%	83.6% 16.4%
	0 0.0%	19 0.1%	0 0.0%	0 0.0%	0 0.0%	2533 13.4%	7 0.0%	0 0.0%	99.0% 1.0%
	0 0.0%	72 0.4%	0 0.0%	0 0.0%	0 0.0%	148 0.8%	2479 13.1%	0 0.0%	91.8% 8.2%
	313 1.7%	31 0.2%	35 0.2%	20 0.1%	6 0.0%	0 0.0%	62 0.3%	2500 13.3%	84.3% 15.7%
Target Class									
	1	2	3	4	5	6	7	8	
	88.5% 11.5%	88.7% 11.3%	64.8% 35.2%	96.5% 3.5%	93.0% 7.0%	94.2% 5.8%	92.0% 8.0%	99.6% 0.4%	91.1% 8.9%

of the last row of the confusion matrix correspond to the Recall values for each class. The percentage at the bottom right cell of the matrix corresponds to the overall accuracy which is the percentage of power signals which have been correctly classified [43].

5 | EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the classification results are presented for the 6-s and 12-s sampling periods of the 24 h signals, using time and frequency domain features separately as well as their combination, utilizing the 1D and 2D-DCT and DFT as described in the previous section. Moreover, the generalization capability of the proposed method is discussed and a comparison of the proposed method with similar home appliance identification methods, is provided.

For the identification process, a FFNN with a single hidden layer is used, where the number of its neurons depends on the length of the feature vector. All process stages and calculations were performed using MATLAB. The results are summarized in Tables 3–6. Specifically, the T largest power values from the time domain and the F largest coefficient values were selected from the frequency domain where $T + F = 5$ for Tables 3 and 5 for a sampling period of 6-s and 12-s, respectively and in Tables 4 and 6 the results are

TABLE 3 (6-s granularity) Classification accuracy scores using five features, the T largest power values and the F largest coefficient values of the DCT or DFT

T	F	Classification accuracy (%)			
		1D-DCT	1D-DFT	2D-DCT	2D-DFT
5	0	90.8	90.8	90.8	90.8
4	1	94.5	94.3	94.4	94.4
3	2	94.6	94.9	95.1	94.8
2	3	93.4	93.6	93.9	93.7
1	4	94.7	92.7	91.6	94.3
0	5	83.9	82.4	83.2	83.3

presented for $T + F = 10$ for a sampling period of 6-s and 12-s, respectively. The selected time and frequency domain features were then introduced to the FFNN for the classification part. In the case where the total number of features used is 5, the hidden layer of the FFNN has 9 neurons while for the case where the total number of features is 10, the number of neurons used is 12. The FFNN architecture was implemented by testing different combinations of input vectors (features) of various sizes with a different number of neurons in the hidden layer each time, utilising empirical rules for artificial neural network design [46].

TABLE 4 (6-s granularity) Classification accuracy scores using 10 features, the T largest power values and the F largest coefficient values of the DCT or DFT

T	F	Classification accuracy (%)			
		1D-DCT	1D-DFT	2D-DCT	2D-DFT
10	0	93.2	93.2	93.2	93.2
8	2	96.1	95.7	96.0	95.8
6	4	95.9	96.3	96.2	96.1
4	6	95.4	95.6	95.6	95.6
2	8	93.5	94.3	93.7	94.1
0	10	86.3	90.3	86.3	86.2

TABLE 5 (12-s granularity) Classification accuracy scores using five features, the T largest power values and the F largest coefficient values of the DCT or DFT

T	F	Classification accuracy (%)			
		1D-DCT	1D-DFT	2D-DCT	2D-DFT
5	0	90.5	90.5	90.5	90.5
4	1	94.7	93.7	93.2	93.2
3	2	95.4	95.5	94.3	95.7
2	3	94.6	94.6	93.5	94.5
1	4	92.1	92.7	92.4	93.1
0	5	83.9	83.0	82.9	82.7

TABLE 6 (12-s granularity) Classification accuracy scores using 10 features, the T largest power values and the F largest coefficient values of the DCT or DFT

T	F	Classification accuracy (%)			
		1D-DCT	1D-DFT	2D-DCT	2D-DFT
10	0	91.1	91.1	91.1	91.1
8	2	96.0	95.7	95.2	96.3
6	4	96.3	96.1	95.7	96.1
4	6	95.7	95.5	95.8	95.9
2	8	94.6	94.8	94.7	95.6
0	10	86.3	89.9	85.7	86.6

In order to confirm the performance and robustness of the proposed method, the 10-cross validation resampling technique was used. All the classification scores presented in this section correspond to the 10 cross-validation mean classification accuracy of the testing set. The proposed method was tested using the appliances listed in Table 2.

From the classification results presented in Tables 3–6 the most important observations could be summarised as follows. For all cases, the classification rates reached when a combination of the time domain (power values) and the frequency domain (coefficients) features are introduced to the FFNN

(rows 2–5 of Tables 3–6) are higher compared to the case where either only the time domain (row 1 of Tables 3–6) or only the frequency domain (row 6 of Tables 3–6) features are used. Another observation is that the corresponding classification rates are similar irrespective of the 6-s or 12-s sampling period indicating that, for the power consumption signals, the information content encapsulated when the signal is sampled at 6 s is similar to the content encapsulated when the same signal is sampled at 12 s. Furthermore, as expected, the use of 10 features (Tables 4 and 6) demonstrate in most of the cases a higher classification rate compared to the corresponding rate when only five features (Tables 3 and 5) are used. However, the improvement, in terms of the classification performance, is not significant, indicating that the time and the frequency domain features selected are compact and encapsulate the information efficiently in their largest values. If more than 10 features are used, the experimental results indicate that they barely improve the classification accuracy score. Closing, the classification results show that for the same granularity and for the same number of frequency domain-only features (row 6 of Tables 3–6) namely, 1D/2D-DFT and DCT coefficients, the recognition performance is similar.

As mentioned earlier, the classification scores demonstrate that a combination of both the time and the frequency domain features always yields better classification scores compared to the case where the features of either the time or the frequency domain are used independently. This improvement occurs because the frequency domain information diversifies the feature set. One example of the aforementioned general observation could be demonstrated through the confusion matrices, where only the time domain (Figure 5 and Table 6 first row), or only the frequency domain (Figure 6 and Table 6 sixth row, 1D-DCT) or a combination of time and frequency domain (1D-DCT) features (Figure 7 and Table 6 third row, 1D-DCT) with 12-s granularity are incorporated for the identification of the home appliances. Specifically, despite the fact that overall the use of only the time domain features provides a higher classification score (91.1%; Figure 5) compared to the case where only the frequency domain features are used (86.3%; Figure 6), for certain classes (appliances) namely: 1 and 8, 3 and 5, 6 and 7, the frequency domain features perform better, in terms of their discrimination capabilities. More importantly, the combination of the time with the frequency domain features provides the highest classification score overall (96.3%, Figure 7) compared to the case where either the frequency or the time domain feature sets are used separately.

In summary, for the experimental results presented until now, the power signals tested were from house 1 of the UK-DALE dataset because house 1 included the domestic appliances which the authors were interested in identifying, as they are used by the majority of the UK households [14]. Moreover, house 1 appliances were monitored for a much longer period of time compared to the appliances of the other four houses of the dataset thus, the number of power signals was much higher, which is important for training the FFNN. In order to further check, though, the generalisation ability of the trained FFNN,

FIGURE 6 Confusion matrix for $T = 0$ and $F = 10$; 12-s granularity (Table 6, sixth row, 1D-DCT column)

Confusion Matrix										
Output Class	1	2689 14.3%	25 0.1%	4 0.0%	0 0.0%	0 0.0%	3 0.0%	84 0.4%	2 0.0%	95.8% 4.2%
	2	1 0.0%	1117 5.9%	221 1.2%	1 0.0%	113 0.6%	0 0.0%	10 0.1%	0 0.0%	76.3% 23.7%
	3	0 0.0%	364 1.9%	1095 5.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	75.1% 24.9%
	4	18 0.1%	89 0.5%	0 0.0%	2051 10.9%	5 0.0%	1 0.0%	595 3.2%	0 0.0%	74.3% 25.7%
	5	0 0.0%	210 1.1%	20 0.1%	0 0.0%	2247 11.9%	0 0.0%	8 0.0%	0 0.0%	90.4% 9.6%
	6	17 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2664 14.1%	85 0.5%	0 0.0%	96.3% 3.7%
	7	92 0.5%	87 0.5%	26 0.1%	322 1.7%	3 0.0%	22 0.1%	1902 10.1%	0 0.0%	77.5% 22.5%
	8	55 0.3%	2 0.0%	47 0.2%	41 0.2%	9 0.0%	0 0.0%	11 0.1%	2508 13.3%	93.8% 6.2%
	93.6% 6.4%	59.0% 41.0%	77.5% 22.5%	84.9% 15.1%	94.5% 5.5%	99.0% 1.0%	70.6% 29.4%	99.9% 0.1%	86.3% 13.7%	
		1	2	3	4	5	6	7	8	
Target Class										

the authors decided to conduct an additional set of experiments using a new testing set consisting of the power consumption signals of the eight appliances from all five houses. Note that for the new testing set all five houses had to be incorporated due to lack of certain classes (e.g. lighting circuit) as well as the limited number of 24 h power consumption signals for some of the other seven classes (appliances) from the other four houses. For this additional set of experiments, the number of power signals reached 380 per appliance (after applying the 3W threshold). The classification rates obtained for this new set of experiments were close to the rates reached using only house 1. As an example, for 12-s granularity, selecting $T = 6$ and $F = 4$ which is the equivalent of Table 6, third row, 1D-DCT column (96.3%), the rate reached using a combination of power signals from all five houses for the eight appliances is 91.7%.

In relation to these results, a useful metric to estimate the generalisation ability of the proposed method is the Generalisation loss (G-loss) metric [47] defined as:

$$G - \text{loss} = 100 \cdot \left(1 - \frac{ACC_u}{ACC_s} \right) \%$$

where ACC_s and ACC_u stand for accuracy on a ‘seen’ house and accuracy on an ‘unseen’ house, respectively. ‘Seen’ is the house whose appliances (power signals) were used to train the

FFNN and test the proposed method in terms of its recognition performance on a ‘known’ dataset and ‘unseen’ is a house whose appliances were used to test the trained FFNN in terms of its generalisation ability. In this case, the 96.3% (0.963) corresponds to the ACC_s and the 91.7% (0.917) to the ACC_u which results in a G-loss of 4.5%, meaning that the classification accuracy on the unseen house is 4.5% lower compared to the seen house. The G-loss would be reduced if power consumption signals from a wider range of home appliance brands and models would be used for the training of the FFNN.

Closing, a comparison of the proposed method, in terms of its recognition performance, with similar published works is presented in Table 7. Comparing different methods of home appliance recognition is not a straightforward task due to the differences in terms of the datasets selected for the experiments, the different sampling periods and lengths of the signals used, the different classes (home appliances) tested etc. [21]. Considering these constraints, a summary of related published research works and a comparison with the proposed method, are provided in Table 7; the common characteristics of the works the proposed method was compared with are as follows: (i) use of real power consumption signals, (ii) use of 24 h power consumption signatures and (iii) similar sampling period (1 s up to 60 s). In all four cases, the recognition performance is similar, however the method proposed here uses a computationally less demanding and easier to implement classifier, as

Confusion Matrix									
Output Class	1	2	3	4	5	6	7	8	
	2694 14.3%	0 0.0%	18 0.1%	11 0.1%	2 0.0%	1 0.0%	56 0.3%	0 0.0%	96.8% 3.2%
	0 0.0%	1801 9.5%	33 0.2%	27 0.1%	8 0.0%	0 0.0%	20 0.1%	1 0.0%	95.3% 4.7%
	0 0.0%	29 0.2%	1299 6.9%	0 0.0%	9 0.0%	0 0.0%	1 0.0%	4 0.0%	96.8% 3.2%
	19 0.1%	7 0.0%	0 0.0%	2321 12.3%	0 0.0%	0 0.0%	57 0.3%	7 0.0%	96.3% 3.7%
	0 0.0%	6 0.0%	28 0.1%	0 0.0%	2352 12.5%	0 0.0%	0 0.0%	0 0.0%	98.6% 1.4%
	3 0.0%	1 0.0%	0 0.0%	1 0.0%	0 0.0%	2675 14.2%	33 0.2%	0 0.0%	98.6% 1.4%
	0 0.0%	25 0.1%	0 0.0%	30 0.2%	0 0.0%	14 0.1%	2520 13.4%	0 0.0%	97.3% 2.7%
	156 0.8%	25 0.1%	35 0.2%	25 0.1%	6 0.0%	0 0.0%	8 0.0%	2498 13.2%	90.7% 9.3%
Target Class									
	1	2	3	4	5	6	7	8	
	93.8% 6.2%	95.1% 4.9%	91.9% 8.1%	96.1% 3.9%	98.9% 1.1%	99.4% 0.6%	93.5% 6.5%	99.5% 0.5%	96.3% 3.7%

FIGURE 7 Confusion matrix for $T = 6$ and $F = 4$; 12-s granularity (Table 6, third row, 1D-DCT column)

TABLE 7 Comparison of the proposed method with related works

	Proposed method	Tundis et al. [32]	Yan et al. [33]	Dash and Naik [26]
Dataset	UK-Dale	Tracebase	Private	UK-Dale
Data signals	24 h power consumption signals sampled at 12 s	24 h power consumption signals sampled at 1 s	24 h power consumption signals sampled at 1 min	24 h power consumption signals sampled at 6 s
Feature set	(i) 10 features (six largest power values and four largest 1D-DCT coefficients; Table 6) (ii) 5 features (three largest power values and two largest 2D-DFT coefficients; Table 5)	19 features extracted from the power consumption signal, the working schedule and the appliance's location	Features extracted from the appliances' feature curve and power range as well as the occupants' behaviour	[200×400] feature matrix extracted from 13 convolution and 3 fully connected layers (CNN)
No. of classes/appliances	8	33	7	8
Classifier	FFNN	Random forest	Bayes model	CNN
Classification accuracy	(i) 96.3% (ii) 95.7%	96.5%	95.6%	97.8%

well as a lower number of features indicating the robustness of the features extracted in terms of encapsulating the power signals' content in an efficient manner. Moreover, non-power consumption-related features such as working schedule, the

appliance's location [32] and the occupant's behaviour [33] are incorporated in the other methods' feature sets; however, these features are not available by the domestic smart metering system which is the focus of this work.

6 | CONCLUSIONS AND FUTURE PLANS

Approximately 21 million smart meters have already been installed in Great Britain. The domestic smart metering system, which is part of the smart cities infrastructure, plays a very important role in the stability of the grid and the introduction of new services such as the ToU tariffs, to mention but a few of the benefits. However, the generated smart metering data can also be utilised for additional energy-related applications such as home appliance identification, thus contributing further to the development and maintenance of the smart city infrastructure.

In this work, a step-by-step description on how consumers can access their smart metering data in real-time using an external CAD via the MQTT API is described. Moreover, an easy-to-implement and computationally efficient method for home appliance identification using power consumption signals which have sampling frequency similar to the sampling frequency available by the UK's domestic smart metering system is implemented and tested. Specifically, features from the time and the frequency domain are introduced to an easy-to-implement FFNN. The experimental results suggest that the recognition rate reaches its highest values, 94%–96%, when features from the frequency domain (1D-DF^T/DCT and 2D-DF^T/DCT) are combined with features from the time domain compared to the case where the time or the frequency domain features are used separately, indicating the complementary nature of the time and frequency domain information. The experimental results also demonstrate the robust compression qualities of the proposed feature extraction method compared to other similar works.

A server-side platform version of the proposed home appliance identification method is currently under development in order to test a variety of computationally demanding time-frequency distributions such as the Wigner-Ville, the Phasegrams [48] and the Continuous Wavelet Transform scalogram, in terms of their compression qualities alongside more advanced classification schemes such as the CNNs.

Furthermore, research is also currently being conducted using a method which is similar to the proposed method in order to identify a household's hot water and space heating consumption. This application is particularly useful for the estimation of a building's heat transfer coefficient (HTC) through the utilisation of the gas consumption data accessed from the domestic smart metering system via an external CAD [49]. HTC estimation using domestic smart metering data is currently an area of particular interest in the field of the built environment.

ORCID

Ioannis Paraskevas  <https://orcid.org/0000-0002-9547-6444>

REFERENCES

- Department for Business, Energy & Industrial Strategy (BEIS): Smart meters: a guide. <https://www.gov.uk/guidance/smart-meters-how-they-work> (2018). Accessed 24 Feb 2021
- Department of Energy & Climate Change (DECC): Smart Metering Implementation Programme. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/245736/smart_meters_domestic_leaflet.pdf (2013). Accessed 24 Feb 2021
- BEIS: Smart Meter Statistics in Great Britain: Quarterly Report to end March 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/887809/Q1_2020_Smart_Meters_Statistics_Report_FINAL.pdf (2020). Accessed 24 Feb 2021
- Loughran, J.: Smart meter rollout delayed to 2024 (Engineering and Technology, IET site). <https://eandt.theiet.org/content/articles/2019/09/smart-meter-rollout-delayed-to-2024/> (2019). Accessed 24 Feb 2021
- DECC: Smart Metering Equipment Technical Specifications 2 (SME^TS2). <https://smartenergycodecompany.co.uk/document-download-centre/download-info/sec-schedule-9-sme-technical-specifications-version-2/> (2018). Accessed 24 Feb 2021
- DCC: SMETS1 Migration - A guide for DCC users. <https://www.smartdccc.co.uk/media/3112/dcc-smets1-installation-migration-guide-for-industry-v10.pdf> (2019). Accessed 24 Feb 2021
- BEIS: Smart Metering Implementation Programme. A report on progress of the realisation of smart meter consumer benefits. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830668/smart-meters-benefits-realisation.pdf (2019). Accessed 9 Feb 2021
- DCC: About DCC. <https://www.smartdccc.co.uk/about> (2021). Accessed 24 Feb 2021
- Morgan, P., Bailey, J.: Capability to support DSR, DECC. https://www.ofgem.gov.uk/sites/default/files/docs/2015/05/20150325_decc_smart_metering_0.pdf (2015). Accessed 24 Feb 2021
- Fitton, R., Swan, W., Yapp, J.: Consumer access to smart data. How do we get consumers connected to their energy data? (BEAMA). <https://www.beama.org.uk/asset/32BD967E-9475-4362-83B553251A84C1F4/> (2018). Accessed 24 Feb 2021
- Manzoni, P.: An IoT oriented comparison, APRICOT 2018: IoT Workshop, Nepal. https://nsrc.org/workshops/2018/apricot/iot/presentations/mqmtsrest_v4.pdf (2018). Accessed 24 Feb 2021
- Chachada, S., Kuo, C.-C.J.: Environmental sound recognition: a survey. *APSIPA Trans. Signal Inf. Pr.* 3, 1–15 (2014)
- UK-DALE dataset, April 2017 release. <https://jack-kelly.com/data/> (2021). Accessed 24 Feb 2021
- Kelly, J., Knottenbelt, W.: The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data.* 2, 1–14 (2015)
- Hart, G.: Nonintrusive appliance load monitoring. *Proc. IEEE.* 80(12), 1870–1891 (1992)
- Zoha, A., et al.: Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. *Sensor.* 12, 16838–16866 (2012)
- Ruano, A., et al.: NILM techniques for intelligent home energy management and ambient assisted living: a review. *Energies.* 12(11), 1–29 (2019)
- de Souza W.A., et al.: Load disaggregation using microscopic power features and pattern recognition. *Energies.* 12, 1–18 (2019)
- Zhao, B., et al.: Non-intrusive load disaggregation solutions for very low-rate smart meter data. *Appl. Energy.* 268, 1–16 (2020)
- Faustine, A., Pereira, L., Klemenjak, C.: Adaptive weighted recurrence graphs for appliance recognition in non-intrusive load monitoring. *IEEE Trans. Smart Grid.* 12, 398–406 (2021)
- Faustine A., Pereira, L.: Multi-label learning for appliance recognition in NILM using Fryze-current decomposition and convolutional neural network. *Energies.* 13, 1–17 (2020)
- Makonin, S., et al.: Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. *IEEE Trans. Smart Grid.* 7(6), 2575–2585 (2016)
- Kim, J.-G., Lee, B.: Appliance classification by power signal analysis based on multi-feature combination multi-layer LSTM. *Energies.* 12(14), 1–24 (2019)
- Kelly, J.: Disaggregation of domestic smart meter energy data. PhD thesis, Imperial College London (2017)

25. Batra, N., et al.: NILMTK: an open source toolkit for non-intrusive load monitoring. Proceedings of the 5th International Conference on Future Energy Systems (ACM e-Energy), Cambridge, UK, July 2014
26. Dash, P., Naik K.: A very deep one dimensional convolutional neural network (VDOCNN) for appliance power signature classification. Proceedings of the 2018 IEEE Electrical Power and Energy Conference (EPEC), Toronto, Ontario, October 2018
27. Penha, D., Garcez Castro, A.R.: Convolutional neural network applied to the identification of residential equipment in nonintrusive load monitoring systems. In: Nagamalai, D., Meghanathan, N. (eds.) AI, CSTY, SIGI – 2017, Computer Science & Information Technology (CS & IT), AIRCC Publishing Corporation, pp. 11–21 (2017)
28. Jiang, J., et al.: Deep learning based energy disaggregation and on/off detection of household appliances. <https://arxiv.org/pdf/1908.00941.pdf> (2019). Accessed 24 Feb 2021
29. Sirojan, T., Phung, B.T., Ambikairajah, E.: Deep neural network based energy disaggregation. Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, Canada, August 2018
30. Çavdar, I.H., Faryad, V.: New design of a supervised energy disaggregation model based on the deep neural network for a smart grid. *Energies*. 12(7), 1–18 (2019)
31. Hamid, O., et al.: Automatic recognition of electric loads analyzing the characteristic parameters of the consumed electric power through a non-intrusive monitoring methodology, International Conference on Technologies and Materials for Renewable Energy, Environment and Sustainability (TMREES), Lebanon, July, 2017
32. Tundis, A., Faizan A., Mühlhäuser M.: A feature-based model for the identification of electrical devices in smart environments. *Sensors*. 19(11), 1–20 (2019)
33. Yan, D., et al.: Household appliance recognition through a Bayes classification model. *Sustain. Cities Soc.* 46, 1–11 (2019)
34. Jin, X., Christensen, D.: Disaggregating smart meter data to identify electric loads and control opportunities. Proceedings of the Non-Intrusive Load Monitoring (NILM), Austin, Texas, March 2018
35. Hopf, K., Sodenkamp, M., Staake T.: Enhancing energy efficiency in the residential sector with smart meter data analytics. *Electron. Markets*. 28(1), 453–473 (2018)
36. Paraskevas, I., et al.: The role of smart meters in energy and cost efficiency in the smart home ecosystem - research in progress. Proceedings of the Living in the Internet of Things (IoT 2019), London, UK, May 2019
37. Hersent, O., Boswarthick, D., Elloumi, O.: The Internet of Things: Key applications and protocols, 2nd ed. John Wiley & Sons (2011)
38. IBM: Introduction to MQTT. https://www.ibm.com/support/knowledgecenter/SSFKSJ_7.5.0/com.ibm.mm.tc.doc/tc00000_h.htm (2021). Accessed 24 Feb 2021
39. European Commission: Order defining the minimum interoperability specifications of the radio transmitter mentioned in Article D. 337-17-1-3 of the Energy Code. <https://ec.europa.eu/growth/tools-databases/tris/en/search/?trisaction=search.detail&year=2016&num=571>, Draft text. (2016). Accessed 24 Feb 2021
40. ZigBee Alliance: Zigbee Smart energy Standard. <https://zigbeealliance.org/wp-content/uploads/2019/11/docs-07-5356-19-0zse-zigbee-smart-energy-profile-specification.pdf> (2014). Accessed 24 Feb 2021
41. Kraken Technologies Ltd.: Octopus energy APIs. <https://developer.octopus.energy/docs/api/> (2021). Accessed 24 Feb 2021
42. Python Package Index (PyPI): Ovoenergy 1.1.11. <https://pypi.org/project/ovoenergy/> (2020). Accessed 24 Feb 2021
43. Theodoridis, S., Koutroumbas, K.: Pattern recognition, 4th ed. Elsevier (2008)
44. Jain, A.K.: Fundamentals of digital image processing, 1st ed. Prentice Hall (1989)
45. Faustine, A., et al.: A survey on non-intrusive load monitoring methods and techniques for energy disaggregation problem. <https://arxiv.org/abs/1703.00785v3> (2017). Accessed 24 Feb 2021
46. Heaton, J.: Introduction to neural networks with Java, chapter. 5, 2nd ed. Heaton Research (2008)
47. Klemenjak, C., et al.: On metrics to assess the transferability of machine learning models in non-intrusive load monitoring. <https://arxiv.org/abs/1912.06200v1> (2019). Accessed 24 Feb 2021
48. Paraskevas, I., Rangoussi, M.: Feature extraction for audio classification of gunshots using the Hartley transform. *Open J. Acoust.* 2(3), 131–142 (2012)
49. Fitton, R.: Energy pathology: measuring a dwelling's energy performance using smart meter and IoT data. *J. Build. Surv. Apprais Valuat.* 9(2), 181–190 (2020)

How to cite this article: Paraskevas I, Barbarosou M, Fitton R, Swan W. Domestic smart metering infrastructure and a method for home appliances identification using low-rate power consumption data. *IET Smart Cities*. 2021;1–16. <https://doi.org/10.1049/smc2.12009>

APPENDIX

```

{
  "elecMtr": {
    "0702": {
      "03": {
        "01": "000001",
        "05": "43",
        "04": "9B",
        "02": "0003E8",
        "07": "",
        "08": "",
        "03": "43",
        "00": "Units: kWh (00)",
        "06": "Commodity: electricity (00)"
      },
      "00": {
        "14": "02",
        "07": "256FB961",
        "01": "000000000000",
        "00": "CurrentSummationDelivered: 1972050 (0000001E1752)",
        "02": "CurrentMaxDemandDelivered: 2422 (000000000976)"
      },
      "04": {
        "01": "001130",
        "00": "0000C1"
      },
      "02": {
        "00": "00"
      }
    }
  },
  "ts": "2019-11-26 10:22:26",
  "hversion": "EHZBRF1v1",
  "time": "5DDCFCE2",
  "gmtime": 1574763746,
  "pan": {
    "status": "joined",
    "nPAN": "00",
    "join": "0",
    "lqi": "FF"
  },
  "gts": "2019-11-26 00:01:56",
  "ets": "2019-11-26 10:22:25",
  "error": {
    "lastCommand": "None (00)",
    "errorCode": "Everything OK - Success (00)"
  },
  "gid": ""
}

```

FIGURE 1A Smart metering cluster/MQTT API (organised form)

TABLE 1A List of abbreviations

Abbreviation	Explanation
API	Application Programming Interface
BEIS	Department for Business, Energy & Industrial Strategy
CAD	Consumer access device
CNN	Convolutional neural network
DCC	Data communications company
DCT	Discrete Cosine Transform
DECC	Department of Energy & Climate Change
DFT	Discrete Fourier Transform
DL	Deep learning
DNN	Deep neural network
ESME	Electricity smart metering equipment
ESP	Energy service portal
FFNN	Feedforward neural network
GSME	Gas smart metering equipment
G-loss	Generalisation loss
HTC	Heat transfer coefficient
IHD	In-home display
LQI	Link quality indicator
MFCC	Mel-frequency cepstral coefficient
MPAN	Meter point administration number
MQTT	MQ telemetry transport
NILM	Non-intrusive load monitoring
PAN	Personal area network
POSIX	Portable operating system interface
REST	Representational state transfer
RNN	Recurrent neural network
SMETS	Smart metering equipment technical specification
ToU	Time of use
UK-DALE	UK domestic appliance-level electricity
1D	One-dimensional
2D	Two-dimensional